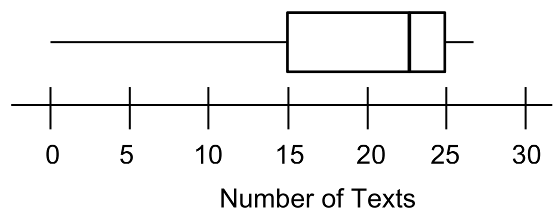
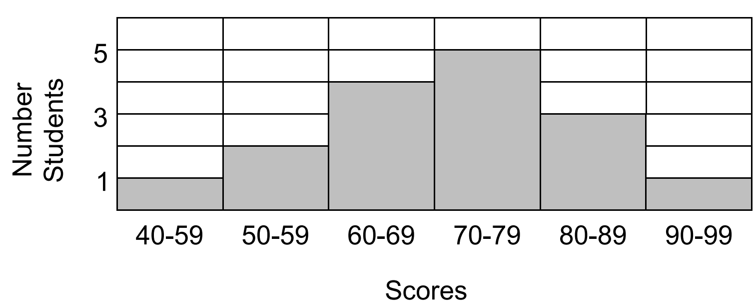
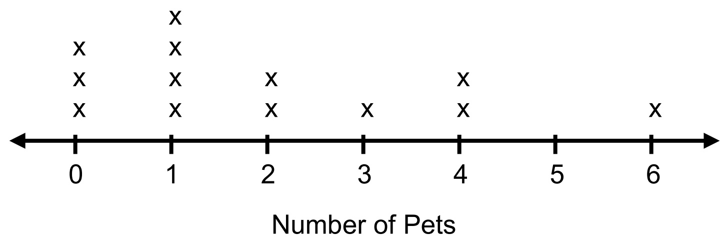
# STUDENT RESOURCES

| Word or Phrase | Definition |
|---|---|
| box plot | A box plot, or box-and-whiskers plot, is a graphical representation of the five-number summary of a data set. See five-number summary.<br><br>Box Plot of Number of Texts Per Day of 6th Graders<br> |
| dot plot | A dot plot is a graphical representation of a data set where the data values are represented by dots above a number line. See line plot. |
| five-number summary | The five-number summary of a data set consists of its minimum value (min), first quartile $Q_1$, median $Q_2$, third quartile $Q_3$, and maximum value (max). The five-number summary is usually written in the form (min, $Q_1$, med., $Q_3$, max).<br><br>The five-number summary of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is given by (min, $Q_1$, med., $Q_3$, max) = (1, 1, 5, 6.5, 23). |
| histogram | A histogram is a graphical representation of frequencies of a numerical variable using rectangles. For a histogram, the horizontal axis is divided into intervals. Each interval forms the base of a rectangle whose height corresponds to the frequency of values of the variable in that interval.<br><br>Quiz Scores of a Class of 16 Students<br> |
| interquartile range | The interquartile range (IQR) of a numerical data set is the difference between the third quartile and the first quartile of the data set. The interquartile range is a measure of the variation of the data set.<br><br>For the data set 1, 1, 1, 3, 5, 5, 6, 7, 23, $Q_1 = 1$, $Q_3 = 6.5$, and IQR = 5.5 |

<table>
<tr><th>Word or Phrase</th><th>Definition</th></tr>
<tr><td>line plot</td><td>A <u>line plot</u> is a graphical representation of a data set where the data values are represented by marks, such as dots or X’s, above a number line. See <u>dot plot</u>.<br><br>Line Plot of Number of Pets for 13 Students<br>
</td></tr>
<tr><td>mean</td><td>The <u>mean</u> of a data set is a measure of center equal to the average of the values in the data set. The mean is calculated by adding the values in the data set and dividing by the number of data values.<br><br>The mean of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is<br><br>$$\frac{1+1+1+3+5+5+6+7+23}{9} = 5\frac{7}{9} = 5.77...$$</td></tr>
<tr><td>mean absolute deviation</td><td>The <u>mean absolute deviation</u> (MAD) of a data set is the average of the (positive) differences between the values in the data set from the mean. The MAD is a measure of the variation of the data set.<br><br>For the data set {3, 3, 5, 6, 6}, the mean is 4.6.<br>The distances of the data points to the mean are 1.6, 1.6, 0.4, 1.4, and 1.4.<br>The MAD is $\frac{1.6+1.6+0.4+1.4+1.4}{5} = 1.28$</td></tr>
<tr><td>measure of center</td><td>A <u>measure of center</u> is a statistic describing the middle of a data set.<br><br>The mean, the median, and the mode are three commonly used measures of center of a numerical data set.</td></tr>
<tr><td>measure of spread</td><td>A <u>measure of spread</u> is a statistic describing the variability of a data set. It describes how far the values in a data set are from the mean or median.<br><br>The standard deviation, the mean absolute deviation (MAD), and the interquartile range (IQR) are three measures of spread of a numerical data set.</td></tr>
<tr><td>median</td><td>The <u>median</u> of a data set is a measure of center equal to the middle number in the data set, when the values are placed in order from least to greatest. If there is an even number of values in the data set, the median is taken to be the mean (average) of the two middle values.<br><br>The median of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is 5, since the first 5 is the middle value.<br><br>The median of the data set 5, 6, 7, 23 is the mean (average) of the two middle numbers, (6 + 7) ÷ 2 = 6.5, which is the average of 6 and 7.</td></tr>
</table>

| Word or Phrase | Definition |
|---|---|
| mode | The mode of a data set is the value(s) that occur(s) most often. A data set may have more than one mode. It may also have no mode if all values occur the same number of times.<br><br>The mode of the data set 1, 1, 1, 3, 5, 6, 6, 7, 23 is 1, since the data value 1 occurs more frequently than any other data value. If a 6 were added to this data set, 6 would also be a mode. |
| outlier | An outlier of a data set is a data value that is a striking deviation from the overall pattern of values in the data set.<br><br>For the data set 1, 1, 1, 3, 5, 6, 6, 7, 23, the data value 23 is a potential outlier. It appears unusually large relative to the other data values. |
| quartiles | The quartiles of a data set are points that divide the data set into four equally sized groups, when the values are placed in order from least to greatest. The second quartile is the median, denoted by $Q_2$. The first quartile, denoted by $Q_1$, is the median of the lower half of the data set (the data values less than the middle data value), and the third quartile, denoted by $Q_3$, is the median of the upper half of the data set.<br><br>Given the ordered data set 1, 1, 1, 3, 5, 5, 6, 7, 23,<br>• The middle value is the first 5: Median = 5. This is also the second quartile $Q_2$,<br>• The lower half of the data set is 1, 1, 1, 3. Therefore $Q_1 = 1$.<br>• The upper half of the data set is 5, 6, 7, 23. Therefore, $Q_3 = 6.5$. |
| range (of a data set) | The range of a numerical data set is the difference between the greatest and least values in the data set.<br><br>The range of the data set 1, 1, 1, 3, 5, 5, 6, 7, 23 is 22, since 22 = 23 – 1. |
| statistical question | A statistical question is a question where numerical data that has potential for variability can be collected and analyzed for the purpose of answering the question.<br><br>A statistical question: “How much TV do students in my class watch on average?”<br>NOT a statistical question: “How many hours of TV did you watch last week?” |

## Finding Measures of Center

Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the median, order the value from least to greatest and find the middle number. If there is an even number of values in the data set, the median is the mean (average) of the two middle numbers.

The median for the siblings data set: 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 5, 5, 7

To find the mode, find the value(s) that occur(s) most often.

The mode for the siblings data set: the value of 2 occurs most often.

To find the mean (average) of a data set, add all the values in the data set and divide it by the number of values (number of observations, $n$).

Number of observations: $n = 13$

The mean for the siblings data set: $\dfrac{3+4+5+2+2+3+3+2+2+5+7+1+1}{13} = 3.08$

## Finding the Range and the Quartiles

Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

To find the range of a data set, find the difference between the greatest and least values in the data set.

For the siblings data set, the range is 6, since 7 – 1 = 6

To find quartiles, first put the numbers in numerical order. Then locate the points that divide the set into four equal parts.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 5 | 5 | 7 |
| ↑ minimum | | | ↑ $Q_1$ | | | ↑ $Q_2$ median | | | | ↑ $Q_3$ | | ↑ maximum |

For the siblings data set:

$Q_1 = 2$ (the 1st quartile)

$Q_2 = 3$ (the 2nd quartile)

$Q_3 = 4.5$ (the 3rd quartile)

Note that $Q_1$ is the median of the first half of the data set and $Q_3$ is the median of the second half.

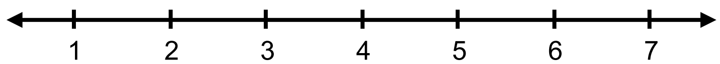
## How to Construct a Dot Plot

A dot plot (also called a line plot) displays data on a number line with a dot ( • ) or an X to show the frequency of data values.

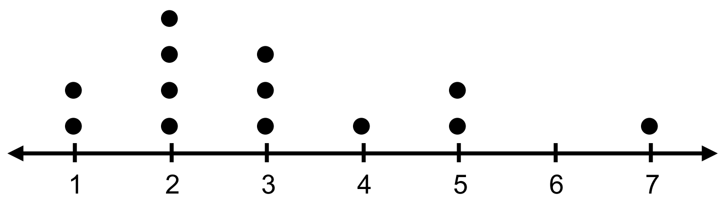
Here are the number of siblings for 13 different students:

3, 4, 5, 2, 2, 3, 3, 2, 2, 5, 7, 1, 1

1. Make a number line that extends from the minimum data value to the maximum data value.

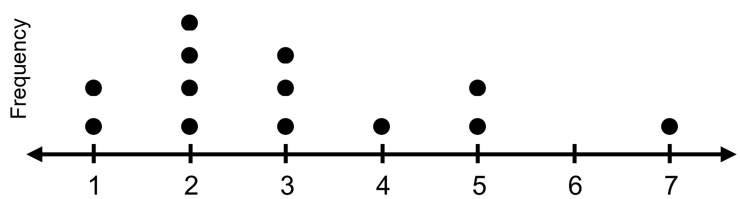



2. Mark a dot or an X for every data value.

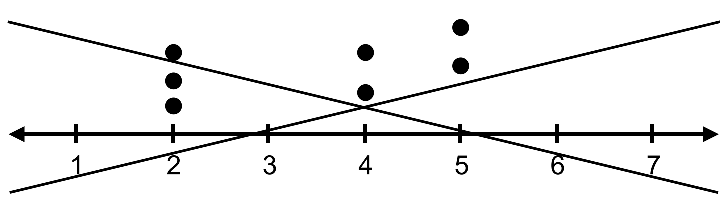



3. Write a title and labels.

Number of Siblings for Students in our Class




**Heads Up!** Be sure to line up dots or X's properly. The graph below is visually misleading in a few places. The number of dots at 4 and 5 are the same, but one set is higher than the other, possibly implying there are more. The number of dots at 2 and 4 are different, but they peak at the same height, possibly implying there are the same number of dots.



## How to Construct a Histogram

A histogram is a data display that uses adjacent rectangles to show the frequency of data values in intervals. The height of a given rectangle shows the frequency of data values in the interval shown at the base of the rectangle.

Nancy asks each of her 21 classmates how many coins they have in their backpacks. Then she puts the data set in order.

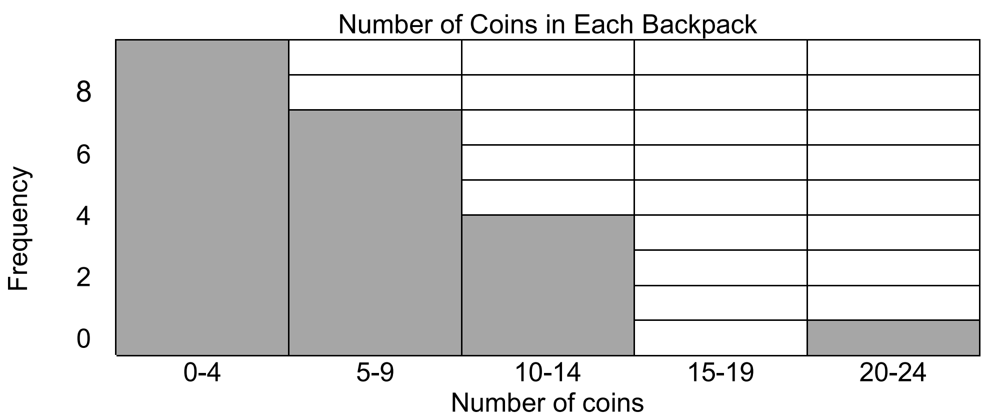
0, 0, 1, 2, 2, 2, 2, 3, 3, 5, 5, 7, 7, 7, 7, 7, 10, 10, 10, 12, 21

To construct the histogram:

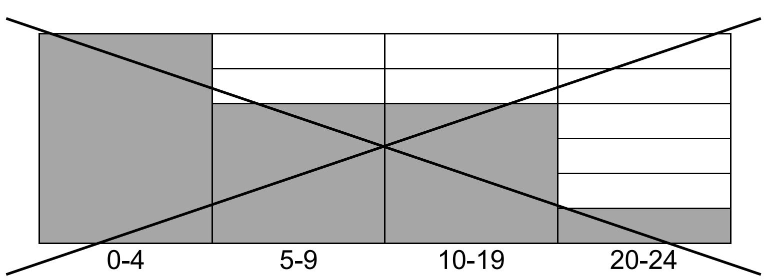
1. Divide the number of coins into equally spaced intervals and make a frequency table: (Here we choose intervals of five.)

| **Intervals (number of coins)** | **Frequency** |
|---|---|
| 0-4 | 9 |
| 5-9 | 7 |
| 10-14 | 4 |
| 15-19 | 0 |
| 20-24 | 1 |

2. Record frequencies as rectangles on a data display. Add a title and label the axes.




**Heads Up!** Be sure to make equally spaced intervals. The graph below is visually misleading. The third column has an interval that is twice the others, but the same number of data points as the column to the left of it



## How to Construct a Box Plot

A box plot (or box-and-whisker plot) is a visual representation of the center and spread of a data set. The display is based on the five-number summary.

Here are the ages of 15 people:

21, 12, 28, 17, 46, 35, 7, 38, 42, 33, 19, 9, 31, 25, 28

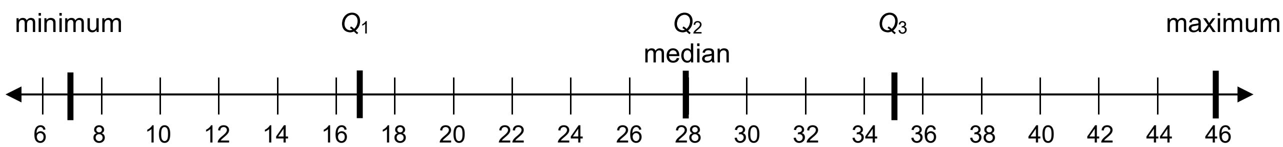
1. Write the values of the data set from least to greatest.

7, 9, 12, 17, 19, 21, 25, 28, 28, 31, 33, 35, 38, 42, 46

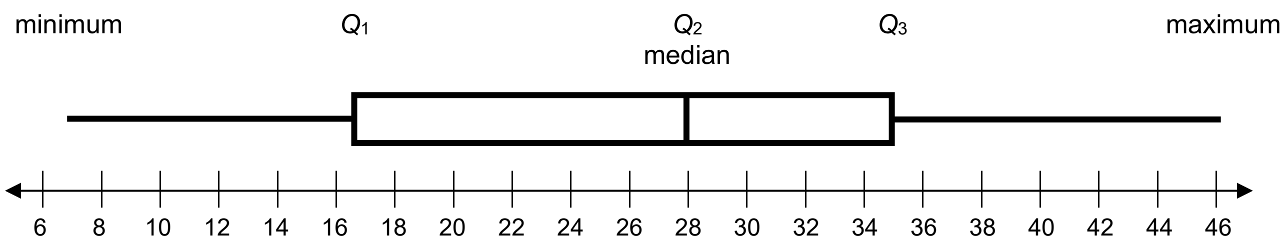
2. Find the five-number summary.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 9 | 12 | 17 | 19 | 21 | 25 | 28 | 28 | 31 | 33 | 35 | 38 | 42 | 46 |
| ↑ minimum | | | ↑ $Q_1$ | | | | ↑ $Q_2$ median | | | | ↑ $Q_3$ | | | ↑ maximum |

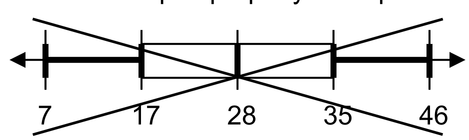
3. Locate the five-number summary values on a number line, and indicate with vertical segments.




4. Create a “box” to highlight the interval from the first to the third quartile, and draw “whiskers” that extend to the minimum and maximum.




**Heads Up!** Be sure to scale the box and whisker plot properly. This plot is WRONG:



# COMMON CORE STATE STANDARDS

| STANDARDS FOR MATHEMATICAL CONTENT | |
|---|---|
| **6.SP.A** | **Develop understanding of statistical variability.** |
| 6.SP.1 | Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers. *For example, "How old am I?" is not a statistical question, but "How old are the students in my school?" is a statistical question because one anticipates variability in students' ages.* |
| 6.SP.2 | Understand that a set of data collected to answer a statistical question has a distribution that can be described by its center, spread, and overall shape. |
| 6.SP.3 | Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number. |
| **6.SP.B** | **Summarize and describe distributions.** |
| 6.SP.4 | Display numerical data in plots on a number line, including dot plots, histograms, and box plots. |
| 6.SP.5 | Summarize numerical data sets in relation to their context, such as by: |
| a. | reporting the number of observations. |
| b. | describing the nature of the attribute under investigation, including how it was measured and its units of measurement. |
| c. | giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered. |
| d. | relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered. |

| STANDARDS FOR MATHEMATICAL PRACTICE | |
|---|---|
| SMP1 | Make sense of problems and persevere in solving them. |
| SMP2 | Reason abstractly and quantitatively. |
| SMP3 | Construct viable arguments and critique the reasoning of others. |
| SMP4 | Model with mathematics. |
| SMP5 | Use appropriate tools strategically. |
| SMP6 | Attend to precision. |
| SMP7 | Look for and make use of structure. |

9781614458371